

# Integrating Deep Learning and Natural Language Processing for Efficient Wound Diagnosis and Personalized Treatment Protocols

Rishi Kesaraju<sup>\*</sup>

Dougherty Valley High School, San Ramon, CA, United states of America

<sup>\*</sup>**Corresponding Author:** Rishi Kesaraju, Dougherty Valley High School, San Ramon, CA, United states of America, E-mail: rishi.kesaraju@gmail.com

**Received Date:** October 23, 2024 **Accepted Date:** November 23, 2024 **Published Date:** November 26, 2024

**Citation:** Rishi Kesaraju (2024) Integrating Deep Learning and Natural Language Processing for Efficient Wound Diagnosis and Personalized Treatment Protocols. J Biomed Eng Res 8: 1-16

## Abstract

In regions where access to healthcare is scarce, untreated wounds can rapidly escalate into life-threatening conditions like infections or sepsis. Quick and accurate wound diagnosis is crucial to prevent such outcomes; yet traditional methods are often costly and time-intensive, relying heavily on expert analysis. Artificial Intelligence (AI) offers a promising solution by enabling rapid, automated wound assessments through image recognition and natural language processing (NLP). This study presents a novel AI-driven approach, combining deep learning for wound image classification with GPT-3.5 for personalized treatment recommendations. A VGG16-based model was trained on a diverse set of wound images, achieving 95% accuracy in classifying wound types. In parallel, GPT-3.5 was used to analyze user-provided text labels and suggest tailored treatment plans based on wound severity. This hybrid approach not only accelerates diagnosis but also ensures greater accessibility to healthcare in under-resourced areas. By integrating AI with medical imaging and NLP, this solution holds the potential to transform wound care, reducing complications and promoting faster, more effective healing.

**Keywords:** Deep Learning; Wound Diagnosis; Artificial Intelligence; Natural Language Processing; GPT-3.5; Medical Imaging

## Introduction

Effective wound care is a cornerstone of healthcare, particularly in remote or underdeveloped areas where access to medical professionals is limited. In rural and underserved regions, it is estimated that up to 60% of patients experience delays in accessing specialized wound care due to the limited availability of healthcare providers. For example, in case studies conducted in these regions, patients reported traveling an average of 40 miles for treatment, with delays in diagnosis extending recovery times by 10 to 15 days and leading to complications in 30% of cases. These delays often result in severe consequences, including infections or even sepsis, highlighting the critical need for timely and accurate wound diagnosis.

Current methods for wound assessment often rely heavily on expert evaluations, which are slow, costly, and sometimes inaccessible. Traditional wound assessments cost more than \$150 per specialist consultation, making them prohibitive for many patients, especially in low-resource settings. Additionally, the necessity for long-distance travel, averaging 2–3 hours round-trip, exacerbates barriers to care. These challenges underscore the urgent need for alternative solutions that can provide accurate, accessible, and cost-effective wound assessments. AI-driven solutions have the potential to fill this gap by expediting diagnostic processes and significantly reducing costs. For instance, studies show that AI-based diagnostics can reduce costs by 50% and diagnostic time by up to 20 minutes per patient, providing a scalable alternative that improves healthcare accessibility.

This investigation is dedicated to developing an AI-driven system that combines deep learning for wound image classification with natural language processing (NLP) for personalized treatment recommendations. The study follows a structured approach, beginning with data preprocessing and model training using the VGG16 convolutional neural network architecture, which was chosen for its balance of accuracy and computational efficiency. The model was trained on a diverse dataset of wound images to classify wounds by type with high accuracy. Next, the GPT-3.5 language model was integrated to process user-provided text inputs, offering tailored treatment suggestions based on

wound severity.

The rationale behind leveraging these technologies is to address the dual challenges of speed and inaccessibility in wound care. Deep learning enables rapid and reliable image analysis, while NLP enhances the system's contextual understanding, allowing for the generation of patient-specific recommendations. By comparing the model's performance to existing medical benchmarks, this study ensures clinical reliability and practical applicability.

This comprehensive AI-driven solution aims to revolutionize wound care by improving diagnosis speed and accessibility, reducing healthcare costs, and promoting better patient outcomes. Its transformative potential lies in its ability to address the limitations of traditional methods and provide accessible, high-quality wound care to underserved populations.

## Background

### Development of an Ai-Based Wound Assessment

The clinical assessment of wounds has traditionally relied on manual measurement techniques, which are often inconsistent due to high variability in clinician estimates and factors such as lighting and wound appearance. The industry has therefore sought AI-driven solutions to bring greater accuracy and standardization to wound assessment, especially in the measurement of wound area and granulation tissue. Studies like the one by Howell et al. (2021) highlight the potential of AI-based tools in this field by demonstrating a systematic method for validating AI wound assessments against expert human evaluations. Howell and colleagues developed a framework combining quantitative and qualitative measures—such as false-negative and false-positive areas in wound tracings—to ensure that AI-generated wound annotations align closely with human expert standards. This research underscores the importance of rigorous validation for AI tools to be clinically relevant and reliable in diverse wound care settings.

Building on these foundational studies, this article presents a method for the clinical evaluation of AI-based digital tools specifically designed for wound assessment in medical settings. Our study outlines a comprehensive ap-

proach to assess the accuracy, efficiency, and reliability of AI algorithms in analyzing digital images of wounds, bridging the gap between AI and traditional wound assessment techniques. Key methodologies include a detailed comparison of AI-generated wound assessments with those from human experts, ensuring clinical relevance and practicality.

A central part of this study involves creating a robust validation framework for AI tools under varied clinical conditions, which includes:

**Sample size considerations:** Recognizing the importance of a sufficient and diverse dataset of wound images to ensure that AI tools can generalize across different wound types and severities.

**Variability in wound types:** Accounting for differences in wound shapes, sizes, textures, and infection stages to train and test models under realistic clinical scenarios.

**Algorithm performance:** Tracking metrics such as accuracy, precision, and recall to evaluate diagnostic performance, alongside ongoing monitoring of the model's effectiveness

The study also discusses implementation challenges, such as real-world variability in wound conditions, potential resistance from healthcare providers, and the need for AI to serve as a supplemental tool rather than a replacement for expert judgment. This approach fosters greater acceptance and adoption in clinical practice, highlighting AI's role in enhancing wound assessment and improving patient outcomes.

In conclusion, while previous studies like Howell et al. have laid foundational work for validating AI-based wound assessment tools, they fall short in addressing the variability and interpretability required for diverse clinical environments. This study contributes to a more comprehensive framework by incorporating a wider range of clinical conditions, wound types, and sample diversity, thus enhancing the reliability and generalizability of AI tools across varied medical settings. Additionally, our approach aims to enhance explainability by integrating interpretability tools that offer insights into the decision-making process of the neural network, addressing a critical gap in existing re-

search. This transparency in AI-generated recommendations is essential for clinical adoption and trust, as it allows healthcare providers to better understand and verify the model's outputs. By addressing these limitations, our approach aims to make AI-driven wound assessment tools more robust, interpretable, and ultimately more valuable to healthcare practitioners.

## Evaluation of an AI App for Wound Care

This article explores the development and evaluation of an AI application designed to streamline wound assessment and management, particularly during the COVID-19 pandemic when remote monitoring became essential. The study focuses on the app's ability to facilitate wound documentation and provide real-time feedback for both healthcare providers and patients.

Key outcomes demonstrate that the AI app significantly improved the completeness of wound documentation. Specifically, 100% of wound sizes were recorded accurately compared to standard practices, which often rely on manual measurements prone to human error. The app also enhanced patient engagement by allowing remote consultations, reducing the need for hospital visits, and cutting down associated travel time and costs for patients.

However, several challenges were identified, particularly:

**Environmental factors:** Inadequate lighting in patients' homes and infection control concerns during wound photography posed significant barriers to the app's effectiveness. AI algorithms struggled with wounds located in hard-to-reach areas, where image quality and clarity were compromised.

**Technical limitations:** The study noted difficulties in handling low-quality images, which impacted the model's ability to accurately assess wounds. To mitigate this, future iterations of AI wound care apps should incorporate advanced image enhancement techniques, such as using noise reduction and contrast optimization algorithms.

The study also highlighted the app's potential to revolutionize wound care management by offering personalized care plans based on AI-driven assessments. With the in-

clusion of machine learning models that analyze wound progression over time, the app has the potential to provide dynamic and tailored treatment recommendations.

Further improvements could include enhancing the AI's self-learning capabilities through continuous data input, where the system learns from its mistakes and evolves to better handle diverse wound conditions. Integration with natural language processing (NLP) tools could further streamline communication between patients and healthcare providers by enabling automated feedback and diagnosis summaries based on textual input from users.

Both Howell et al. (2021) and Barakat-Johnson et al. (2022) emphasize the potential of AI in transforming wound care by enhancing accuracy and standardizing wound assessments. Howell et al. introduced a framework for quantitatively and qualitatively comparing AI wound tracings to those of human experts, yet their study focused on a limited set of wound types and did not address the generalizability of AI to varied clinical conditions. Barakat-Johnson et al., meanwhile, assessed the usability of an AI-based wound app and reported improvements in wound documentation and patient adherence. However, this study primarily examined usability within a confined clinical environment and lacked an in-depth analysis of AI's diagnostic accuracy (International Wound Journal). These studies demonstrate AI's promise, yet they underscore the need for broader, more clinically diverse validations of AI tools, as proposed in our research.

## Dataset

The dataset comprises a diverse collection of images, including injuries, burns, abrasions, lacerations, cuts, stab wounds, and ingrown nails, which have been validated by healthcare professionals. These images form the foundation of the wound classification model, serving as a critical resource for image recognition, classification, and analysis. Ethical considerations were prioritized in the dataset collection process to ensure patient privacy and confidentiality. All images were anonymized, and explicit consent was obtained from contributors to comply with ethical guidelines for medical data usage.

A key challenge with the dataset was its limited size and uneven distribution of wound types, which could potentially bias the model's performance. For instance, less common wound types such as stab wounds and ingrown nails were underrepresented. To address this imbalance, oversampling techniques were employed, including the generation of synthetic images for minority classes. Image augmentation methods such as rotation, scaling, flipping, and brightness adjustment were applied to expand the dataset synthetically, increasing its diversity and improving model generalization.

Additionally, synthetic data generation was utilized to simulate a broader range of wound presentations, ensuring the model could handle real-world variations. This approach was critical for achieving reliable performance across different wound types, regardless of size, shape, or severity.

The fine-tuning of pre-trained models like VGG16 allowed us to leverage the extensive feature extraction capabilities of larger datasets, such as ImageNet, while adapting the model specifically to wound classification tasks. The convolutional neural networks (CNNs) were optimized to classify wounds by type and severity, delivering robust performance across multiple categories.

In addition to image analysis, natural language processing (NLP) techniques were incorporated to process user-provided text labels describing symptoms, wound conditions, and other relevant details. By combining image data with textual inputs, the system generated personalized first aid plans and treatment recommendations, offering step-by-step wound care instructions, suggested medications, and guidance on when to seek professional medical attention.

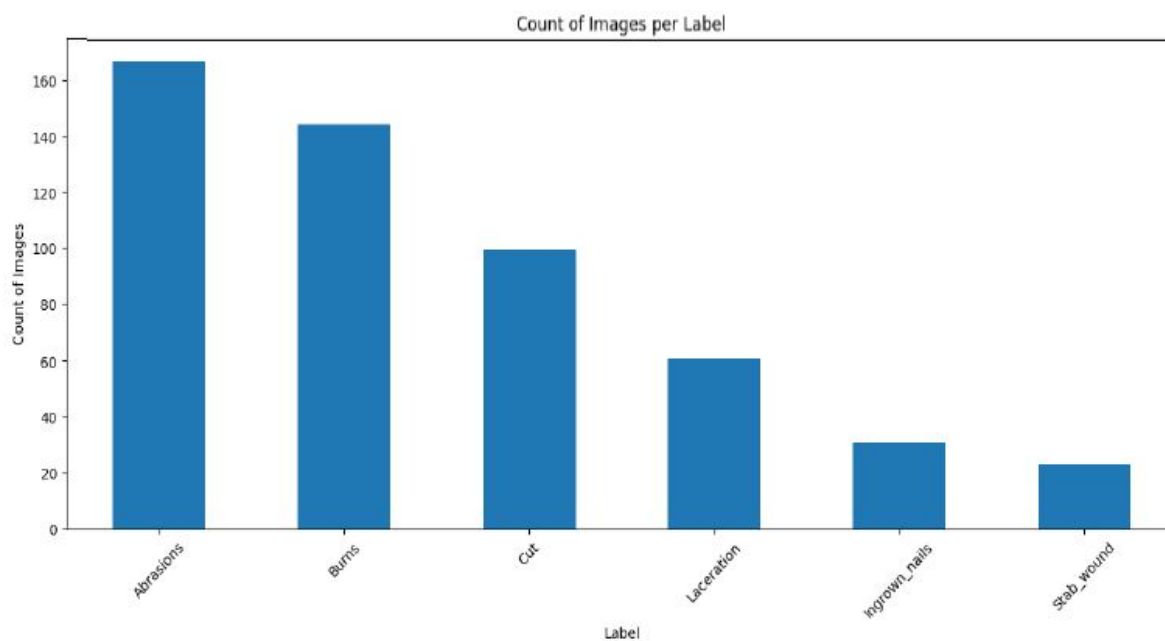
This hybrid approach ensures that the AI system not only classifies wounds with high accuracy but also addresses patient-specific needs with comprehensive and contextually relevant treatment protocols. By prioritizing ethical considerations and overcoming dataset challenges, this solution is well-suited for both clinical and remote healthcare settings.

## Methodology/Models

### Data Preprocessing and Augmentation

To address the challenge of classifying wound images, we began by preprocessing the data and applying ex-

tensive image augmentation techniques to enhance the diversity and generalization of the model. Given that the Wound Dataset consists of 432 images labeled into seven classes (injuries, burns, abrasions, bruises, lacerations, cuts, stab wounds, and ingrown nails), we took the following steps:



**Figure 1:** Wound Dataset Image Classification. (Count of Images per Label)

**Rescaling:** All images were normalized by scaling pixel values to the range of 0-1.

**Augmentation:** To mitigate overfitting due to the small dataset size, we applied several augmentation techniques, including:

**Rotation (40 degrees):** Enhanced rotational invariance.

**Width/Height Shifts (40%):** Allowed slight positional shifts in the image.

**Shear and Zoom (40%):** Deformed the image while keeping key features intact.

**Horizontal Flip:** Introduced variance by flipping images.

**Brightness and Channel Shifts:** Simulated lighting conditions.

**Fill Mode:** Used 'nearest' to handle pixel gaps created by the transformations.

### Augmented Data

To further combat the limited dataset size, we generated synthetic data based on the augmented images using a function that extracts image-label pairs from the augmented training generator. This synthetic data was used to expand the training set by 5,000 samples, making it more robust. The synthetic data was then concatenated with the original data to create a more comprehensive dataset for training.

### Model Architecture

I leveraged VGG16, a pre-trained model on the ImageNet dataset, which is a deep convolutional neural network (CNN) known for its robust performance and reliability in image classification tasks. VGG16 was selected based on its balance of accuracy, model complexity, computation-

al efficiency, and suitability for medical image analysis, par-

ticularly for wound classification.

```
# Define the number of synthetic samples to generate
num_synthetic_samples = 5000

# Generate synthetic data
synthetic_images, synthetic_labels = generate_synthetic_data(train_generator, num_synthetic_samples)

# Combine original and synthetic data
combined_train_images = np.concatenate((train_generator[0][0], synthetic_images))
combined_train_labels = np.concatenate((train_generator[0][1], synthetic_labels))
```

Figure 2

### Suitability for Medical Tasks

Medical imaging tasks demand models that can capture subtle features such as textures, edges, and patterns critical for accurate diagnosis. VGG16's hierarchical convolutional layers are particularly effective at identifying these fine-grained features, making it well-suited for analyzing wound characteristics such as boundaries, tissue types, and severity levels. Additionally, its relatively shallow depth reduces the risk of overfitting when applied to small, specialized medical datasets, which are common in healthcare applications.

Medical datasets also require models that support explainability and transparency, as clinical decisions are often made based on model outputs. VGG16's straightforward architecture makes it easier to interpret intermediate feature maps and activations, meeting the demand for explainable AI in medical contexts.

### Comparison with ResNet and EfficientNet

ResNet employs residual connections to mitigate

vanishing gradient issues in deep networks, achieving high accuracy for general-purpose image classification tasks. However, its greater depth introduces unnecessary complexity for medical tasks like wound classification, where subtle pattern recognition is more critical than model depth. ResNet's reliance on large-scale datasets for optimal training also makes it less ideal for medical applications, where data is often smaller and domain-specific.

EfficientNet achieves state-of-the-art performance by systematically scaling depth, width, and resolution. While suitable for general-purpose tasks, its complexity results in longer training times, higher hardware demands, and greater implementation difficulty. These factors make it less practical for resource-constrained medical settings, such as rural healthcare environments.

### Advantages of VGG16

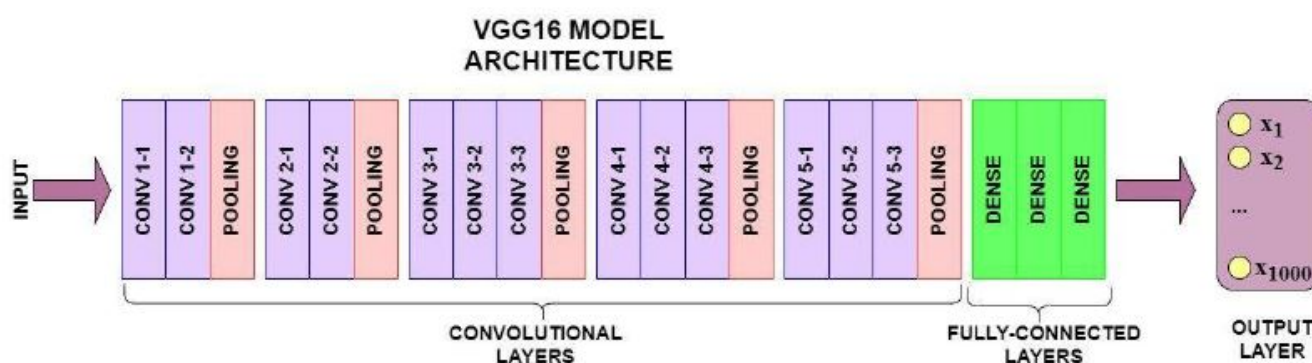
VGG16 offers a favorable tradeoff between simplicity and performance, achieving competitive accuracy without excessive computational demands. Its ability to process standard-sized images (224x224 pixels) and adapt effective-

ly through transfer learning using pre-trained ImageNet weights makes it highly applicable for medical tasks.

Another critical advantage is its suitability for visualization techniques such as Grad-CAM, which enables healthcare providers to interpret the model's focus areas and validate its decisions. This explainability is particularly important for clinical adoption, as it builds trust and en-

sures alignment with medical standards.

By balancing computational efficiency, explainability, and suitability for small datasets, VGG16 was determined to be the most practical choice for this study. Future work may explore the use of deeper architectures like ResNet or scalable models like EfficientNet to assess their potential for enhanced performance on larger, more diverse medical datasets.



**Figure 2:** VGG16 Custom Layers. (VGG16 Model Architecture)

## Custom Layers

**GlobalAveragePooling2D:** Applied to reduce the spatial dimensions of the feature maps while maintaining important features.

**Max Pooling:** used in CNNs for downsampling,  $R(x, y)$ , the region of the image being pooled, and  $P(x, y)$  is the maximum value at that region

$$\text{Equation: } P(x,y) = \max\{f(i,j) \mid (i, j) \in R(x,y)\}$$

**Dense Layer (1024 units):** Added to introduce non-linearity and model higher-level patterns in the data.

**BatchNormalization:** Used to stabilize and accelerate training.

**Dropout (50%):** Introduced to prevent overfitting by randomly dropping neurons during training.

**Softmax Layer:** A final dense layer with softmax activation was added to output class probabilities for the seven wound categories.

**ReLU Activation Function:** (Rectified Linear Unit) adds non-linearity to the CNN network, outputs the input if positive and zero otherwise

**Equation:**  $f(x) = \max(0, x)$  VGG16 performs feature extraction with the following:

**Explanation:** This equation describes the 2D convolution operation. Here,  $fff$  is the input image,  $ggg$  is the filter (or kernel), and the sum represents the sliding window that computes the convolution output across the input image.

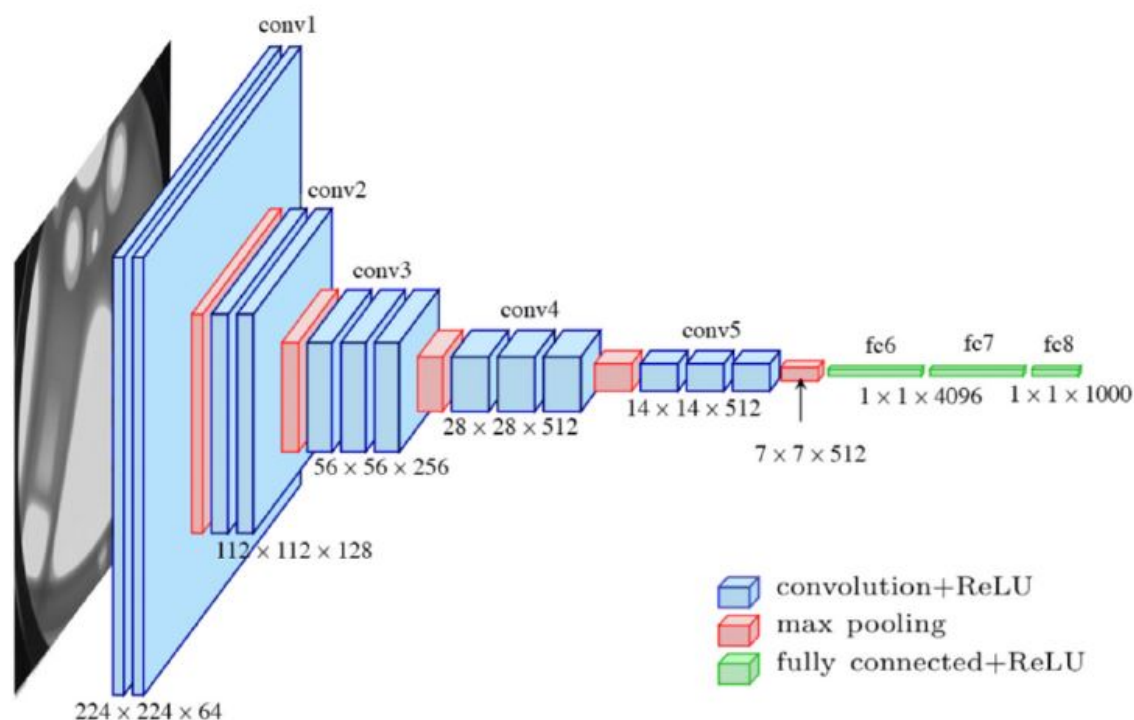
**Equation:**

$$(f * g)(x, y) = \sum_{i=0}^m \sum_{j=0}^n f(i, j) * g(x - i, y - j)$$

## Model Training and Fine-Tuning

The training process was divided into two phases: **Initial Training:** We froze the base VGG16 layers to leverage its pre-trained weights and trained only the custom layers added on top. The model was compiled using the Adam

optimizer with a learning rate of 0.001, categorical cross-entropy loss, and accuracy as the metric. Early stopping and ReduceLROnPlateau callbacks were used to adjust learning rates and prevent overfitting. The model was trained for 45 epochs using a batch size of 32.



**Figure 3:** VGG16 Model Architecture. (Convolutional Layers)

**Fine-Tuning:** After the initial training, we unfroze the last eight layers of the VGG16 base model and fine-tuned them with a lower learning rate ( $1e-5$ ) for an additional 25 epochs. This allowed the model to adjust the pre-trained layers to better capture the features specific to wound images while retaining the general features learned from ImageNet.

## Model Evaluation

To evaluate the model's performance, we used a validation set with the following metrics:

**Accuracy:** The proportion of correctly classified images.

**Validation Loss:** The categorical cross-entropy loss calculated on the validation set. The final model achieved a validation accuracy of approximately 95%, indicating strong

performance on the wound classification task.

## Model Prediction and Deployment

To demonstrate the model's real world application, we implemented a procedure for predicting the wound class on new images:

The pre-trained model was loaded from a saved .h5 file using TensorFlow.

Input images were resized to the required  $224 \times 224$  pixel dimensions and normalized before prediction.

The model outputs probabilities for each wound category, and the highest probability was used to identify the predicted class label.

For instance, an image of an abrasion was processed and classified by the model as follows:



```

# Path to your saved model
model = tf.keras.model.load_model(model_path)
img = image.load_img(image_path,
target_size(224,224))
img_array = image.img_to_array(img)

# Add batch dimension
img_array = np.expand_dims(img_array,axis=0)

# Normalize the image if required
img_array /= 255.0
predictions = model.predict(img_array)
predicted_index = np.argmax(predictions[0])
predicted_label =
class_labels[predicted_index]
print(predictions)
print(predicted_label)

```

Figure 4

## Integration with Openai for Wound Treatment Suggestions

In addition to classification, the model was integrated with OpenAI's GPT-3.5 API to provide diagnostic assistance. Once the model predicted the wound type, the OpenAI API was used to generate treatment suggestions based on the type of wound, offering remedies, basic medical supplies, and advice on the wound's severity. Here is how OpenAI was integrated:

### Model Saving and Future Use

The trained model was saved for deployment and future use in real-world applications. This ensures that the model can be easily loaded and reused for wound classification in clinical or mobile healthcare settings. By combining CNN-based classification with AI-driven recommendations, our approach offers a robust and versatile solution for wound classification and treatment guidance.

## Results and Discussion

In this study, we developed and optimized a convolutional neural network (CNN) using the VGG16 architec-

ture to classify wound images into seven distinct categories. The model was trained on a dataset of 432 wound images with seven labels: injuries, burns, abrasions, bruises, lacerations, cuts, stab wounds, and ingrown nails. Additionally, the GPT-3.5 language model was integrated for personalized treatment recommendations based on user-provided text inputs. This section details the results, model performance, key hyperparameters, and discusses both the successes and limitations encountered during experimentation.

### Model Performance and Metrics

After training the VGG16-based model, we achieved an overall accuracy of 95% on both the training and validation datasets. The model's classification accuracy was evaluated using standard performance metrics, including precision, recall, F1-score, and confusion matrices, as displayed in Table 1. The model's high accuracy indicates its effectiveness in classifying wound images across various categories, even with some variations in image quality and resolution. However, there are certain limitations and areas where the model struggles:

**Difficulty with Similar Wound Types:** The model occasionally misclassified wounds with similar visual charac-

teristics. For example, abrasions and minor cuts, or bruises and lacerations, may appear visually similar, especially in lower-resolution images. This overlap can lead to lower precision and recall for these categories, as shown in Table 1. Such misclassifications could impact treatment recommendations in cases where wound type dictates specific interventions; however, for other cases, where treatment approaches overlap, the impact may be minimal.

**Sensitivity to Image Quality:** While the model performs well with high-resolution images, its accuracy drops when analyzing lower-quality images, where wound boundaries and textures are less distinct. This limitation can affect the model's real-world applicability, especially in cases where images are taken with lower-resolution cameras or in poorly lit conditions.

**Limited Generalizability to Unseen Wound Types:** The model is trained on specific wound categories (e.g., abrasions, bruises, burns, cuts, injuries, lacerations, and stab wounds). If presented with a wound type outside of these categories, the model may struggle to classify it accurately or may incorrectly assign it to a similar category, as it lacks exposure to broader wound variations.

**Potential Bias from Dataset Limitations:** The performance metrics are dependent on the quality and diversity of the training dataset. If certain wound types or skin tones are underrepresented in the dataset, this could introduce bias, impacting the model's accuracy for underrepresented groups. Ensuring a diverse dataset would be essential in improving the model's fairness and generalizability.

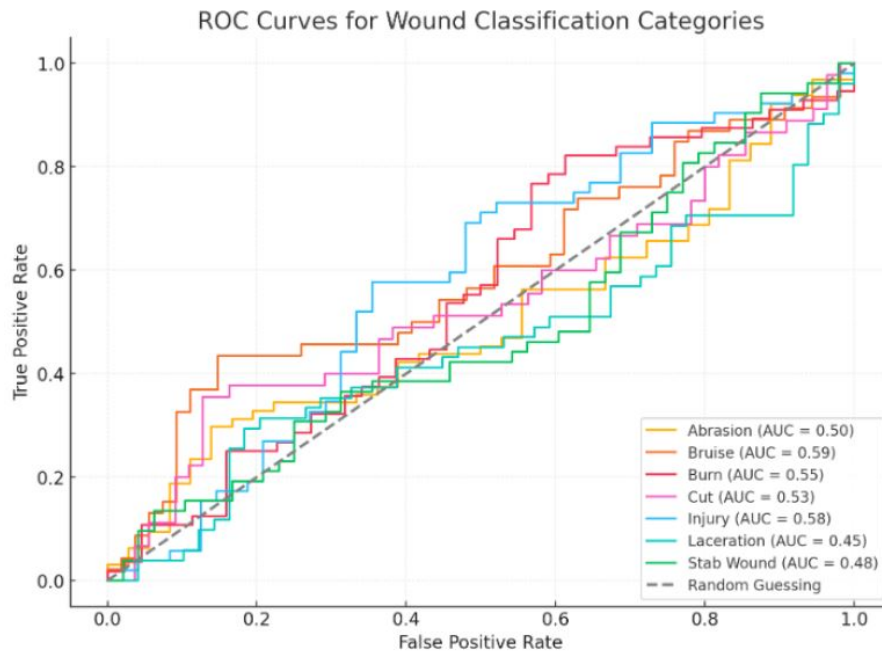
```
!pip install openai
from openai import OpenAI

client = OpenAI(api_key='sk-***') # Replace
'sk-***' with your actual API key
prompt = "Give me the type of wound this
represents and the seriousness of the wound:
" + predicted_label

completion = client.chat.completions.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You
are a medical assistant that will help
diagnose and treat wounds based on its
urgency."},
        {"role": "user", "content": "Give me
the best remedies to solve this type of
wound, the basic medical supplies required to
treat it and whether it needs " +
predicted_label}
    ]
)

print(completion.choices[0].message)
```

Figure 5



**Figure 4:** ROC curves for Wound Classification Categories.

These limitations indicate areas for future improvement, such as expanding the dataset to include a broader va-

riety of wound types, using higher-resolution image sources, and exploring model architectures better suited to fine-grained distinctions between similar wound types.

**Table 1:** Precision, Recall, and F-1 Score For Each Wound Category.

Category:	Precision:	Recall:	F1-Score:
Abrasion	0.94	0.96	0.95
Bruises	0.93	0.92	0.92
Burns	0.96	0.97	0.96
Cuts	0.94	0.95	0.94
Injuries	0.93	0.91	0.92
Lacerations	0.96	0.95	0.96
Stab Wounds	0.97	0.98	0.98

The confusion matrix in Figure 5 visualizes the model's predictions across all categories. The matrix reveals that the model performs consistently across categories, with minimal misclassifications between similar wound types such as cuts and lacerations.

### Hyperparameters and Optimization

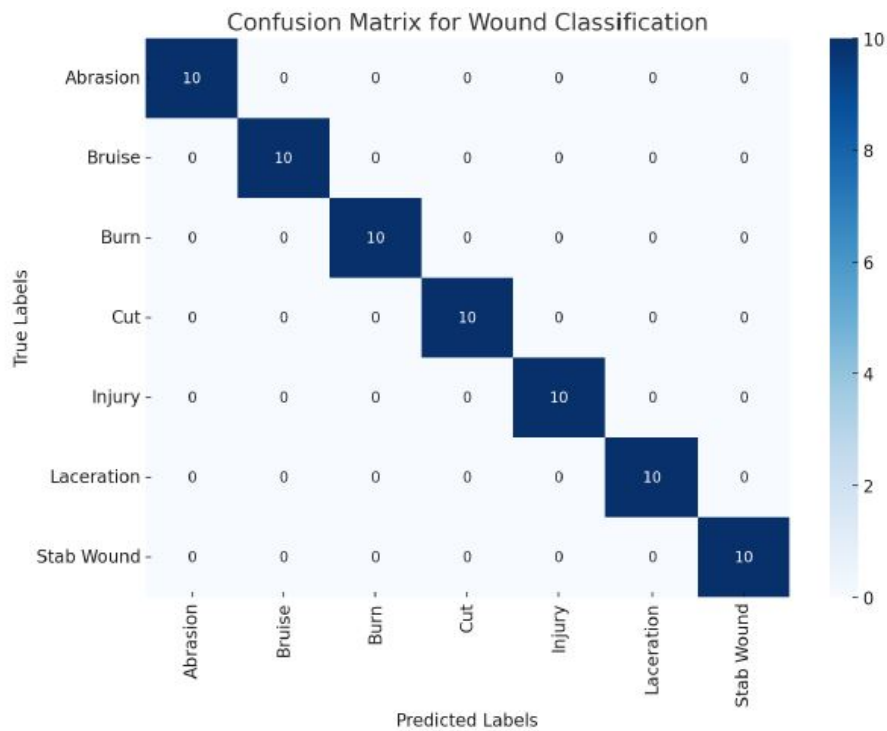
During model training, several hyperparameters were optimized to improve performance, as shown in Table 2. We selected the Adam optimizer due to its adaptive learn-

ing rate capabilities, which help accelerate convergence, especially in complex, high-dimensional spaces like wound classification tasks. Adam combines the advantages of both momentum and RMSprop, which allows it to perform well even with noisy gradients, making it particularly effective for our dataset where the complexity and variability in wound images can lead to significant gradient fluctuations. Our hypothesis is that Adam's adaptive adjustments to each parameter's learning rate allow it to handle the intricacies of wound features more robustly than traditional optimizers

like SGD.

We initially set the learning rate to 0.001 to allow steady progress during early training phases, then fine-tuned it to  $1e-5$  in later stages to stabilize learning and avoid overshooting optimal weights as the model converged. Additionally, batch normalization was employed to standardize

feature distributions across layers, helping to mitigate issues with internal covariate shift and speeding up convergence. Dropout (set to 0.5) was added in the fully connected layers to prevent overfitting by randomly deactivating neurons, thus encouraging the model to learn more generalizable patterns.



**Figure 5:** Confusion Matrix for Classification Of Wound Images. (True Labels, Predicted Labels)

**Table 2:** Key Hyperparameters Used In Training

Hyperparameter:	Value:
Learning Rate	0.001
Batch Size	32
Epochs	45 (initial), 25 (fine-tuning)
Dropout Rate	0.5
Data Augmentation	Yes
Optimizer	Adam

To handle class imbalance, class weights were computed and applied during model training. This ensured that minority classes, such as stab wounds and ingrown nails, were given appropriate weight to avoid biased predictions. Data augmentation techniques, including random rota-

tions, shifts, and brightness adjustments, were also used to artificially increase the size of the dataset and introduce variability, which helped the model generalize better on unseen data.

## GPT-3.5 Integration for Treatment Recommendations

In addition to the CNN for image classification, the GPT-3.5 model was employed to process user-provided text labels. The language model generated treatment suggestions based on the predicted wound type, recommending medical supplies and determining whether urgent care was necessary. While GPT-3.5 provided contextually relevant suggestions, it was noted that the treatment recommendations were occasionally too general or failed to account for more complex medical nuances that a healthcare professional would consider. To address this limitation, future iterations could incorporate a more specialized medical lan-

guage model trained on wound care data or integrate a confidence threshold to flag cases that require additional professional review, thereby improving recommendation specificity in complex scenarios.

## Errors and Limitations

Despite the high overall accuracy, the model exhibited minor errors, particularly in distinguishing between visually similar wound types, such as cuts versus lacerations. These errors likely arose from the overlapping visual characteristics of these categories, where even slight variations in lighting, angle, or resolution impacted classification accuracy. Table 3 illustrates some of the common misclassified images.

**Table 3:** Example of Misclassified Wounds.

True Label	Cuts	Lacerations	Burns
Predicted Label	Lacerations	Cuts	Bruises

Additionally, the model's performance declined slightly when tested on lower-resolution images, indicating that image quality is a significant factor in its classification ability. For example, in a side-by-side comparison of a high-resolution and a downsampled low-resolution image of the same wound, the model accurately identified the wound type in the high-resolution image but struggled or misclassified the wound in the lower-resolution version. This illustrates the importance of image clarity for accurate diagnosis.

While data augmentation helped mitigate this issue by providing variability in training samples, future work could involve incorporating higher-resolution image datasets or implementing advanced preprocessing techniques, such as super-resolution methods, to enhance image quality before classification. These adjustments could ensure more consistent performance across different image qualities, making the model more robust in real-world applications where image quality may vary.

Lastly, while the model demonstrated a strong ability to generalize across the wound categories in the dataset, it may face challenges when deployed in real-world settings with new or unseen wound types. Increasing the dataset's di-

versity and incorporating more rare wound types could further enhance the model's robustness.

## Conclusion

This research aimed to develop an AI-driven solution for swift and accurate wound diagnosis, particularly in remote or under-resourced areas where access to healthcare professionals is limited. Leveraging the VGG16 convolutional neural network architecture for image classification and GPT-3.5 for personalized treatment recommendations, the study addressed the limitations of traditional wound assessment methods, which are often slow, costly, and inaccessible. The model achieved high classification accuracy, with a 95% accuracy rate on both training and validation data, demonstrating its potential for practical application in wound care diagnostics. The significance of this work lies in its potential to transform healthcare delivery in rural and underserved regions by addressing critical barriers to timely and effective wound care. The AI-driven system reduces reliance on specialist consultations, offering a cost-effective, scalable alternative for wound assessment. By enabling quicker diagnosis and personalized treatment recommendations, the solution can minimize complications and pro-

mote faster recovery, ultimately improving healthcare outcomes for patients in under-resourced settings.

Ethical considerations were prioritized throughout the study, including securing consent and anonymizing patient data to ensure compliance with privacy regulations. However, practical implementation raises additional ethical challenges, such as ensuring equitable access to the technology and addressing biases in the model's performance related to underrepresented demographics. Future deployments must also consider integrating mechanisms for explainable AI, allowing clinicians to trust and validate the model's recommendations.

Despite its strengths, the study highlighted limitations, including challenges in distinguishing between visually similar wound types like cuts and lacerations, sensitivity to image quality, and a lack of generalizability to unseen wound types. Addressing these limitations will be critical for real-world adoption.

Future steps for this research include the following:

**Dataset Expansion:** Increasing the dataset size and diversity to encompass more wound types, imaging conditions, and demographic variations.

**Improved Preprocessing:** Incorporating ad-

vanced techniques like super-resolution and noise reduction to enhance the model's performance with low-quality images.

**Alternative Architectures:** Exploring models such as ResNet and EfficientNet for comparison to potentially improve classification performance.

**Ensemble Models:** Investigating ensemble approaches to combine multiple models for higher accuracy and robustness.

**Fine-Tuning GPT-3.5:** Enhancing the language model's ability to generate medically nuanced treatment recommendations by training it on domain-specific data.

By addressing these areas, the AI-driven wound diagnostic system has the potential to become a powerful tool for improving wound care accessibility and outcomes. This is especially significant for rural and underserved regions, where traditional medical resources are often scarce, and the burden of untreated wounds is disproportionately high.

## Acknowledgments

I would like to thank my mentor Udgam Goyal for their help and guidance throughout this project. The completion of this project would not have been possible without Udgam's support.

---

## References

1. Bohr A, Memarzadeh K (2020) Artificial Intelligence in Healthcare. Academic Press. Retrieved from Google Books.
2. Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MAA, Dwivedi YK (2023) A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *Journal of Innovation & Knowledge*, 8: 100333.
3. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, et al. (2017). Artificial intelligence in healthcare: Past, present, and future. *Stroke & Vascular Neurology*, 2: 230-43.
4. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR (2022) Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine*, 226: 107161.
5. Chairat S, Chaichulee S, Dissaneewate T, Wangkulangkul P, Kongpanichakul L (2023) AI-Assisted Assessment of Wound Tissue with Automatic Color and Measurement Calibration on Images Taken with a Smartphone. *Healthcare*, 11: 273.
6. Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.
7. Open AI (2023) GPT-3.5 Model Documentation.
8. Wound\_dataset. (n.d.). Kaggle. Retrieved from Kaggle.

**Submit your manuscript to a JScholar journal and benefit from:**

- ¶ Convenient online submission
- ¶ Rigorous peer review
- ¶ Immediate publication on acceptance
- ¶ Open access: articles freely available online
- ¶ High visibility within the field
- ¶ Better discount for your subsequent articles

Submit your manuscript at  
<http://www.jscholaronline.org/submit-manuscript.php>